



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **2000221990 A**(43) Date of publication of application: **11.08.00**

(51) Int. Cl.

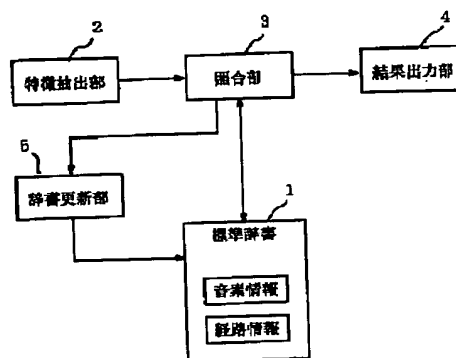
G10L 15/06(21) Application number: **11019555**(22) Date of filing: **28.01.99**(71) Applicant: **RICOH CO LTD**(72) Inventor: **KURODA MASARU**(54) **VOICE RECOGNIZING DEVICE**

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognizing device capable of satisfactorily preparing a dictionary for recognizing a specific speaker without casting a burden on the speaker.

SOLUTION: This voice recognizing device has a standard dictionary 1, a feature extracting part 2 for extracting the quantity of features from inputted voice, a collating part 3 for collating the quantity of features extracted by the feature extracting part 2 with the standard dictionary 1, a result output part 4 for outputting results of the collation in the collating part 3, and a dictionary renewing part 5 for renewing the standard dictionary 1. The standard dictionary 1 is prepared, in its initial stage, as a dictionary for recognizing unspecific speakers, and the dictionary renewing part 5 is adapted to renew the standard dictionary 1 into a dictionary for recognizing a specific speaker based on results of the collation of the quantity of features extracted from inputted voice with the standard dictionary 1.

COPYRIGHT: (C)2000,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-221990
(P2000-221990A)

(43) 公開日 平成12年8月11日 (2000.8.11)

(51) Int.Cl.⁷
G 1 0 L 15/06

識別記号

F I
G 1 0 L 3/00

テーマコード(参考)

5 2 1 S 5 D 0 1 5

審査請求 未請求 請求項の数 4 O L (全 5 頁)

(21) 出願番号 特願平11-19555

(22) 出願日 平成11年1月28日 (1999.1.28)

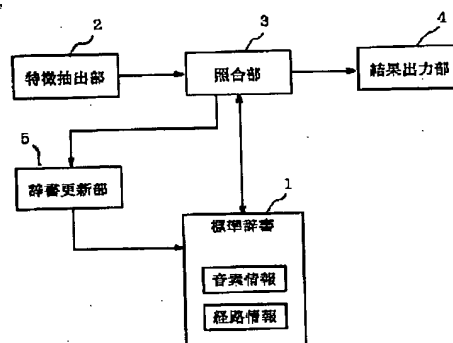
(71) 出願人 000006747
株式会社リコー
東京都大田区中馬込1丁目3番6号
(72) 発明者 黒田 勝
東京都大田区中馬込1丁目3番6号 株式
会社リコー内
(74) 代理人 100090240
弁理士 植本 雅治
Fターム(参考) 5D015 AA02 GG01

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 特定話者認識用の辞書を発声者に負担をかけずに、かつ、良好に作成することの可能な音声認識装置を提供する。

【解決手段】 この音声認識装置は、標準辞書1と、入力された音声から特徴量を抽出する特徴抽出部2と、特徴抽出部2で抽出された入力音声の特徴量と標準辞書1との照合を行なう照合部3と、照合部3における照合結果を出力する結果出力部4と、標準辞書1の更新を行なう辞書更新部5とを有している。ここで、標準辞書1は、初期時においては、不特定話者認識用の辞書として作成されており、辞書更新部5は、入力音声から抽出された特徴量と標準辞書1との照合結果に基づいて標準辞書1を特定話者認識用の辞書に更新するようになっている。



【特許請求の範囲】

【請求項 1】 標準辞書と、入力された音声から特徴量を抽出する特徴抽出部と、特徴抽出部で抽出された入力音声の特徴量と標準辞書との照合を行なう照合部と、照合部における照合結果を出力する結果出力部と、標準辞書の更新を行なう辞書更新部とを有し、前記標準辞書は、初期時においては、不特定話者認識用の辞書として作成されており、前記辞書更新部は、入力音声から抽出された特徴量と標準辞書との照合結果に基づいて標準辞書を特定話者認識用の辞書に更新するようになってい

10 ことを特徴とする音声認識装置。

【請求項 2】 請求項 1 記載の音声認識装置において、前記標準辞書は、初期時において、文字列から音素情報を抽出して不特定話者認識用の辞書として作成されており、前記照合部は、ある文字列についての入力音声から抽出された特徴量により決められる入力音素と標準辞書の前記文字列に対応する音素情報との照合において、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離評価を行ない、前記辞書更新部は、前記音素距離評価結果に基づいて、標準辞書の前記文字列の対応する音素情報を更新し、標準辞書を特定話者認識用の辞書に更新することを特徴とする音声認識装置。

20 【請求項 3】 請求項 2 記載の音声認識装置において、前記辞書更新部は、前記音素距離評価の結果、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書の前記文字列の対応する音素情報を更新し、標準辞書を更新することを特徴とする音声認識装置。

30 【請求項 4】 請求項 2 記載の音声認識装置において、前記辞書更新部は、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離評価を行ない、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書の前記文字列の母音に対応する音素を更新し、標準辞書を更新することを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識装置に関する。

【0002】

【従来の技術】一般に、音声認識装置において、特定話者認識を行なうためには、特定話者に所定の単語等について発声を行なわせて特定話者用の辞書の作成を行っていた。この場合、特定話者用の辞書を作成するためには、特定話者は、1つの単語等について1回から3回発声する必要がある。

【0003】ここで、1回しか発声しない場合は、発声者の負担は比較的少なくて済むが、辞書登録時の環境（例えば、背景の騒音や、周囲の人の発声）によって必ずしも良い辞書ができない。これは、特定話者が発声した

音声に周囲の音が混入してしまうために、登録された辞書の品質が下がってしまうためである。

【0004】これに対し、特定話者に複数回（例えば3回）の発声を行なわせて辞書の作成（登録）を行なう場合には、例えば、複数回の発声に基づいて平均的な辞書を作成することができる。あるいは、1回目の発声で辞書を作成し、2回目あるいは3回目の発声で辞書との照合を行なって辞書の出来を評価することもできる。いずれにしても、1回しか発声しない場合に比べて良い辞書ができる。

【0005】

【発明が解決しようとする課題】しかしながら、発声者に同じ単語等を2回あるいは3回繰り返して発声させる場合、発声者にとって辞書作成が負担となってしまう。例えば、20単語も30単語も辞書登録することになると、非常に負担となる。

【0006】本発明は、特定話者認識用の辞書を発声者に負担をかけずに、かつ、良好に作成することの可能な音声認識装置を提供することを目的としている。

【0007】

【課題を解決するための手段】上記目的を達成するために、請求項 1 記載の発明は、標準辞書と、入力された音声から特徴量を抽出する特徴抽出部と、特徴抽出部で抽出された入力音声の特徴量と標準辞書との照合を行なう照合部と、照合部における照合結果を出力する結果出力部と、標準辞書の更新を行なう辞書更新部とを有し、標準辞書は、初期時においては、不特定話者認識用の辞書として作成されており、辞書更新部は、入力音声から抽出された特徴量と標準辞書との照合結果に基づいて標準辞書を特定話者認識用の辞書に更新するようになってい

40 ることを特徴としている。

【0008】また、請求項 2 記載の発明は、請求項 1 記載の音声認識装置において、前記標準辞書は、初期時において、文字列から音素情報を抽出して不特定話者認識用の辞書として作成されており、前記照合部は、ある文字列についての入力音声から抽出された特徴量により決められる入力音素と標準辞書の前記文字列に対応する音素情報との照合において、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離評価を行ない、前記辞書更新部は、前記音素距離評価結果に基づいて、標準辞書の前記文字列の対応する音素情報を更新し、標準辞書を特定話者認識用の辞書に更新することを特徴としている。

【0009】また、請求項 3 記載の発明は、請求項 2 記載の音声認識装置において、前記辞書更新部は、前記音素距離評価の結果、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書の前記文字列の対応する音素情報を更新し、標準辞書を更新することを特徴としている。

50 【0010】また、請求項 4 記載の発明は、請求項 2 記

載の音声認識装置において、前記辞書更新部は、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離評価を行ない、入力音素と標準辞書の前記文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書の前記文字列の母音に対応する音素を更新し、標準辞書を更新することを特徴としている。

【0011】

【発明の実施の形態】以下、本発明の実施形態を図面に基づいて説明する。図1は本発明に係る音声認識装置の構成例を示す図である。図1を参照すると、この音声認識装置は、標準辞書1と、入力された音声から特徴量を抽出する特徴抽出部2と、特徴抽出部2で抽出された入力音声の特徴量と標準辞書1との照合を行なう照合部3と、照合部3における照合結果を出力する結果出力部4と、標準辞書1の更新を行なう辞書更新部5とを有している。ここで、標準辞書1は、初期時においては、不特定話者認識用の辞書として作成されており、辞書更新部5は、入力音声から抽出された特徴量と標準辞書1との照合結果に基づいて標準辞書1を特定話者認識用の辞書に更新するようになっている。

【0012】より具体的に、標準辞書1は、初期時において、文字列から音素情報を抽出して不特定話者認識用の辞書として作成されており、照合部3は、ある文字列についての入力音声から抽出された特徴量により決められる入力音素と標準辞書1の文字列に対応する音素情報との照合において、入力音素と標準辞書1の文字列に対応する音素情報との音素距離評価を行ない、辞書更新部5は、音素距離評価結果に基づいて、標準辞書1の文字列の対応する音素情報を更新し、標準辞書1を特定話者認識用の辞書に更新するようになっている。

【0013】ここで、標準辞書1は、初期時において、文字列から音素情報、経路情報が抽出されて、不特定話者認識用の辞書として作成されている。例えば、「あき」という文字列について標準辞書1を作成したい場合には、この文字列「あき」を、「a」「a-k」「k」「k-i」「i」というように音素単位に分ける。これら4個の音素を標準辞書1の中に音素情報として保存する。また、経路情報も同様に「a」「a-k」「k」「k-i」「i」という順序で遷移していくことを標準辞書1に保存する。なお、この文字列から音素情報、経路情報への変換は、音声認識装置内で行なっても良いし、別の文字列変換装置を使用して行なっても良い。

【0014】この場合、入力された音声例えば「あき」であるとする、照合部3は、入力音声「あき」を解析した結果の経路情報「a」「a-k」「k」「k-i」「i」の順序と一致する経路情報が標準辞書1に存在するかを調べ、いまの場合、「a」「a-k」「k」「k-i」「i」の順序と一致する経路情報が標準辞書1に存在するので、入力音声「あき」の音素(入力音素)「a」「a-k」「k」「k-i」「i」と上記経路情

報が一致した標準辞書1の音素情報「a」「a-k」「k」「k-i」「i」との音素距離を算出するようになっている。すなわち、この場合、音素距離Xは、例えば、入力音声「あき」の音素「a」「a-k」「k」「k-i」「i」と標準辞書1の音素「a」「a-k」「k」「k-i」「i」との各々対応する音素の距離X(「a」), X(「a-k」), X(「k」), X(「k-i」), X(「i」)の加算値として算出される。

【0015】そして、辞書更新部5は、音素距離評価の結果、入力音素と標準辞書1の文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書1の文字列の対応する音素情報を更新し、標準辞書1を更新するようになっている。すなわち、上記の例では、標準辞書1に現在登録されている音素情報「a」「a-k」「k」「k-i」「i」を入力音声の音素情報「a」「a-k」「k」「k-i」「i」に差し替えるようになっている。

【0016】あるいは、辞書更新部5は、入力音素と標準辞書1の文字列に対応する音素情報との音素距離評価を行ない、入力音素と標準辞書1の文字列に対応する音素情報との音素距離が所定の閾値を越えたときにのみ、標準辞書1の文字列の母音に対応する音素を更新し、標準辞書1を更新することもできる。すなわち、上記の例では、標準辞書1に現在登録されている音素情報「a」「a-k」「k」「k-i」「i」のうち、母音に対応する音素「a」、「i」を、入力音声の音素情報「a」「a-k」「k」「k-i」「i」中の母音に対応する音素「a」、「i」に差し替えることもできる。

【0017】なお、このような標準辞書1の更新処理は、標準辞書1に登録するための発声となされる場合になされても良いし、音声認識のための発声となされる場合になされても良い。すなわち、入力音声は、認識のための発声であっても良いし、登録のための発声であっても良い。

【0018】次に、図1に示す構成の音声認識装置の動作について説明する。まず、標準辞書1に登録するための発声となされる場合について説明する。話者は、ある文字列について標準辞書1に登録のための音声を1回発声し、これが装置内に入力されると、入力された音声は、特徴抽出部2で音声特徴量に変換される。

【0019】特徴抽出部2で変換された音声特徴量は、照合部3において経路探索、類似度演算がなされる。ここで、経路探索の過程において、音素と経路の遷移情報により、図2に示すように、入力された音声、例えば「あき」の特徴量の音素(入力音素)と、標準辞書1の音素情報との対応がとれる。入力音素と標準辞書の音素情報との対応がとれると、入力音素と標準辞書の対応する音素情報との音素距離Xが求められる(音素距離評価がなされる)。このように求められた距離は一時的に保存する。このようにして、照合部3において、入力音素と

標準辞書 1 の音素情報との照合が終了した後、辞書更新部 5 は、下記手順に従って、標準辞書 1 の音素情報を更新する。

【0020】すなわち、辞書更新部 5 は、まず、入力音素と標準辞書 1 の音素情報との音素距離 X が所定の閾値を越えたか否かを判断する。この結果、入力音素と標準辞書 1 の音素情報との音素距離 X が所定の閾値を越えない場合には、標準辞書 1 の音素情報は、特定話者の入力音素に類似し、特定話者の音声の特徴を良好に反映しているとみなせるので、標準辞書 1 の音素情報は更新しない。

【0021】これに対し、入力音素と標準辞書 1 の音素情報との音素距離 X が所定の閾値を越えた場合には、標準辞書 1 の音素情報は、特定話者の音声の特徴を良好に反映しているとはみなせないで、標準辞書 1 の音素情報を入力音素に入れ替える。

【0022】また、このとき、母音のみを更新することもできる。母音のみを更新するときは、文字列情報によりどの音素が母音であるかを判定し、母音と判定された音素に、母音であることを表わすタグを付け、そのタグがある音素について、上述したと同様の処理を行なう。

【0023】次に、音声認識のための発声がなされる場合について説明する。音声認識のための発声がなされる場合においては、登録のための発声がなされる場合の上述した手順に、さらに、以下の手順が加わる。すなわち、音声認識のための発声がなされる場合には、音声認識装置が出力した認識結果を発声者が確認する必要がある。すなわち、登録のための発声であれば対象となる単語は発声する前に決められている（これから発声する単語は何であるか発声者も音声認識装置もわかっている）ので、その単語と発声を上述の手順通りに実行すればよいが、音声認識のための発声時の場合は、たとえ照合類似度が大きな値となったとしても音声認識装置が誤認識する可能性がある。そこで発声者が認識出力結果を確認したうえで上述の手順を実行する必要がある。

【0024】このように、本発明では、標準辞書 1 は、初期時においては、不特定話者認識用の辞書として作成されており、辞書更新部 5 は、入力音声から抽出された特徴量と標準辞書 1 との照合結果に基づいて標準辞書 1 を更新するようになっているので、標準辞書 1 を更新して特定話者認識用の辞書とするのに、発声者は 1 回の発

声だけで済み、特定話者認識用の辞書を発声者に負担をかけずに、かつ、良好に作成することができる。

【0025】また、本発明では、認識結果（類似度）によって辞書全体を更新するのではなく、辞書の中の音素のみを更新するようになっているので、これにより、音声認識装置全体の構成の中で、揮発メモリに蓄えておく容量が少なくて済む。

【0026】

【発明の効果】以上に説明したように、請求項 1 乃至請求項 4 記載の発明によれば、標準辞書と、入力された音声から特徴量を抽出する特徴抽出部と、特徴抽出部で抽出された入力音声の特徴量と標準辞書との照合を行なう照合部と、照合部における照合結果を出力する結果出力部と、標準辞書の更新を行なう辞書更新部とを有し、標準辞書は、初期時においては、不特定話者認識用の辞書として作成されており、辞書更新部は、入力音声から抽出された特徴量と標準辞書との照合結果に基づいて標準辞書を特定話者認識用の辞書に更新するようになっているので、特定話者認識用の辞書の作成に必要な発声者の負担を著しく軽減することができ、さらに、不特定話者認識用の辞書の情報を使用して、特定話者認識用の辞書に近い認識辞書が作成できるために、高性能でかつユーザインタフェースの優れた音声認識装置を提供することができる。

【0027】特に、請求項 4 記載の発明では、文字列の母音に対応する音素のみを更新するので、更新する情報が少ない割には、更新されたことによる辞書性能を著しく向上させることができ、音声認識装置の構成を小さく、かつ高性能にすることができる。

【図面の簡単な説明】

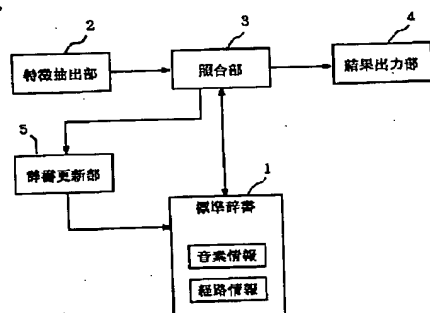
【図 1】本発明に係る音声認識装置の構成例を示す図である。

【図 2】入力された音声、例えば「あき」の特徴量の音素（入力音素）と、標準辞書の音素情報との対応をとる様子の一例を示す図である。

【符号の説明】

1	標準辞書
2	特徴抽出部
3	照合部
4	結果出力部
5	辞書更新部

【図 1】



【図 2】

